SAND 2000-8439

ASCI WAN at SC99

A Step on the Path to 100 Gigabit Per Second

Leveraging on the advances made by the National Transparent Optical Network (NTON) dense wave division multiplexing (DWDM) infrastructure, we prototype a leading edge wide area network (WAN) to meet the bandwidth demand of large-scale ASCI applications. This testbed uses state-of-the-art packet-over-SONET (POS) links of the Avici terabit switch routers (TSR) to connect Sandia/CA, LBNL, JPL, and the ASCI booth. Local area network (LAN) resources are accessed via Packet Engines' PowerRail 5200 at Gigabit Ethernet speeds. To demonstrate the performance benefit of this testbed, we run the Image Based Rendering Assisted Volume Rendering (IBRAVR) application developed at LBNL. This application loads data from a Distributed Parallel Storage System (DPSS) at LBNL, performs parallel 2D-image compositing on SNL/CA Cplant, and assembles 3D images for rendering at the ASCI booth.

The NTON POS Testbed

The diagram below depicts the NTON POS testbed. As shown, the core of the testbed consists of two Avici TSRs, one at Oakland and the other at Los Angeles. These routers are connected via OC48c (2.5 Gbps) links multiplexed from NTON's OC192 (10 Gbps) infrastructure. From NTON's Oakland point of presence (POP), two OC12c POS links are used to connect SNL/CA's Avici TSR. Using a proprietary protocol, the Avici TSR builds a composite trunk to deliver twice the OC12c rate (1.2 Gbps). Also connected to the Oakland TSR is an OC12c link to LBNL, where additional traffic from other research testbeds is expected. At the LA POP, a DWDM OC48c connection is used to reach the Portland POP. From there, the OC48c link is extended to the show floor via Nortel's Optera hardware. Through this link, the Avici TSR at the ASCI booth connects to the POS testbed at an OC48c rate.

Testbed Equipment

At the SNL/CA location, we allocated 32

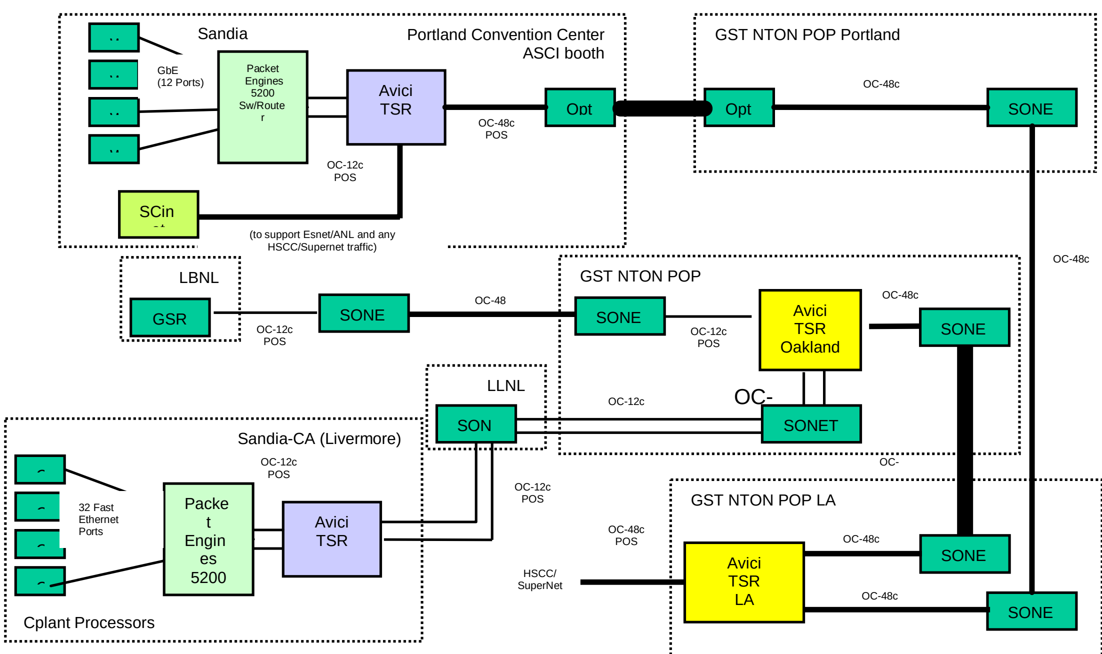

Figure 1. SC99 POS Network

Alpha processors from our Cplant for this demonstration (Figure 2). These compute nodes are connected to the POS testbed indirectly through a Packet Engines PowerRail 5200 switch router, which has an aggregate bandwidth of 52 Gbps and can support up to 64 Gigabit Ethernet connections. It is the only Gigabit Ethernet switch that offers an OC12c POS interface and we use it to interconnect our Cplant nodes and to reach the POS testbed.

The Avici router is designed to meet the demands of the exponential growth in the Internet. The Avici TSR architecture can scale along both the bit-rate and port-count axes. Today, the Avici router supports POS ports ranging from OC3c, OC12c, and OC48c, with OC192c planned for the first quarter of 2000. The TSR switch fabric adapts a direct connection network architecture, where each network line card is also a switching node, all of which are connected to form a 3D torus, using six 40 Gbps fabric links. This architecture allows incremental growth to achieve 364 Tbps of aggregate backplane bandwidth that can support tens of thousands of Gigabit Ethernet ports.

High performance File Server

At LBNL, we use the Distributed Parallel Storage System data block server to provide high-performance data handling for our demonstration application. The DPSS architecture builds high-performance storage systems from low-cost, commodity hardware components. This technology has been quite successful in providing an economical, high-performance, widely distributed, and highly scaleable architecture for caching large amounts of data that can potentially be used by many different users.

Remote Visualization Demo

Over this testbed, we demonstrate a high-resolution, interactive remote visualization technique developed at LBNL under the NGI Combustion Corridor project. From the 32 Cplant nodes at SNL/CA, the Image Based Rendering Assisted Volume Rendering (IBRAVR) application loads raw volume data of a combustion calculation from the DPSS at LBNL. This operation uses the Message Passing Interface (MPI) library to achieve high I/O rates through parallelism. Upon data-load completion, the 32 Cplant nodes perform 2D-image compositing in parallel, each using a small subset of the overall data volume to achieve scalability. Resulting 2D images are transmitted, using parallel TCP/IP sockets, over the NTON testbed to a rendering engine at the ASCI or the LBNL booth. The rendering engines in turn assemble the received 2D images into 3D representations for display either on the ASCI Powerwall or the VR hardware at the LBNL booth.

Summary

The DOE ASCI project is chartered with the simulation of science and engineering problems of unprecedented size and complexity, often requiring the use of human and compute resources that are distributed at geographically distributed locations. As such, very high bandwidth networks as well as innovative remote visualization software tools are essential to ensure the success of the ASCI program. This prototype testbed explores leading network technologies available to the ASCI communities. We run the LBNL IBRAVR application to validate these technologies because it demands high speed accesses of data, compute, and visualization resources at distributed locations.

For more information contact:

H. Chen, J. Brandt, and P. Wyckoff
Sandia National Laboratories
MS 9011, PO Box 969
Livermore, CA 94551-0969

N. Johnston et al.
Lawrence Berkeley National Lab
One Cyclotron Road, MS 50F University of Berkeley
Berkeley, CA 94720

http://www.llnl.gov/NTON/
http://www.cs.sandia.gov/cplant/
http://www-didc.lbl.gov/DPSS/Overview/DPSS.handout.fm.html
http://rockford.lbl.gov/projects/ibravr/

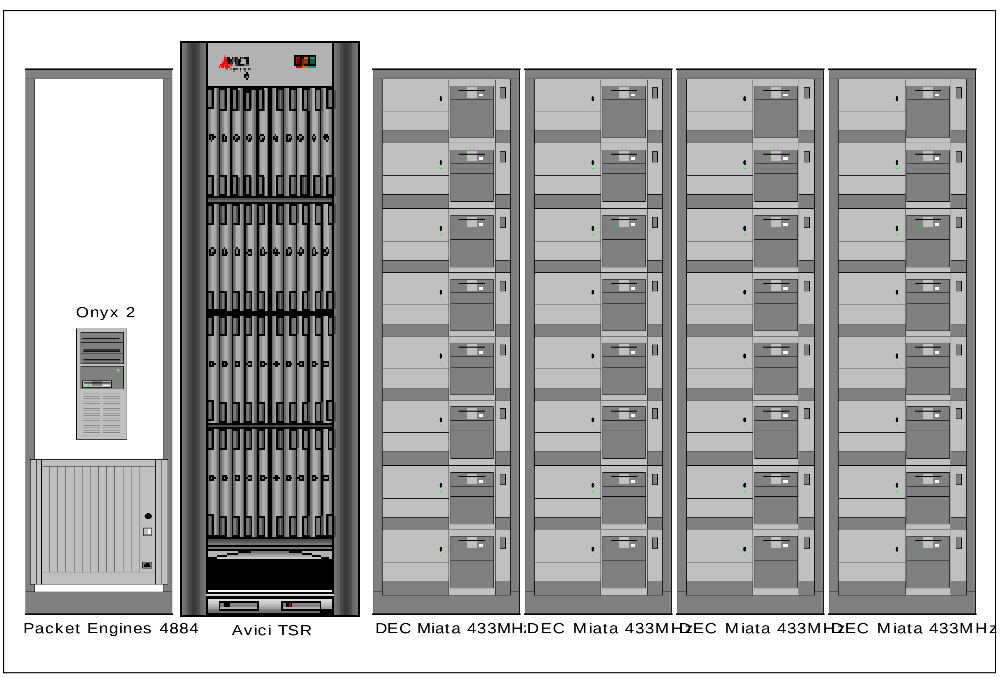

Figure 2. SNL/CA testbed equipment.